GEOPHYSICAL RESEARCH LETTERS, VOL. 27, NO. 20, PAGES 3333-3336, OCTOBER 15, 2000

The lognormal distribution as a reference for reporting aerosol optical depth statistics; Empirical tests using multi-year, multi-site AERONET sunphotometer data

O'Neill, N. T.[1,3], Ignatov, A.[2], Holben, B. N.[1], Eck, T. F.[1]

Abstract Aerosol optical depth data representative of various types of aerosols was employed to empirically demonstrate that the lognormal probability distribution is a better reference for reporting optical depth statistics than a normal probability distribution.

Introduction

Measurements of natural variables are often better characterized by probability distributions which more closely resemble a lognormal (ℓ) distribution than a normal (m) distribution (see Campbell, 1995 for example). Aerosol optical depth (AOD) measurements are regularily reported in terms of the basic statistical parameters associated with the n distribution (arithmetic mean and standard deviation) rather than the basic parameters associated with the ℓ distribution (geometric mean and geometric standard deviation). Some authors have indicated that the latter representation may be more appropriate (Malm et al., [1977], King and Byrne [1980], Ignatov and Ratner [1995], Ignatov and Stowe [19991) and accordingly that AOD histograms and associated parameters should be referenced to log τ space rather than τ space.

Users of AOD statistics require that the reported parameters adequately mimic sample histograms so that derived quantities meet the accuracy needs of model driven applications such as radiative forcing or aerosol dispersion. The more complex the parameterization the more freedom one has to achieve better histogram characterizations. However, the increased complexity necessarily renders all associated operations more difficult and ultimately wasteful when the level of parameterization is excessive relative to the information content of the measurements. It is thus of some importance to search for a degree of parameterization which is as simple as possible while achieving a level of distribution characterization which is commensurate with application requirements.

In this note we empirically evaluate the applicability of arithmetic and geometric parameters in log τ and τ space and of the associated ℓ and n probability distributions to a multi-year and multi-site AOD database in order to demonstrate that the ℓ distribution is systematically superior to the n distribution within the constraint of a uni-modal probability distribution.

The normal and lognormal probability distributions

The probability distribution of a series of N measurements in x can be written;

$$P(x) = \frac{1}{N}\frac{dN}{dx} \qquad (1)$$

where dN is the number of samples in the increment dx and N is the total number of samples. The use of the exact derivative in equation (1) implies that the series of N measurements has been repeated an arbitrarily large number of times. The ℓ distribution is simply a normal distribution with x = log τ;

$$P_\ell(\log\tau) = \frac{1}{\sqrt{2\pi}\,\log\mu}\exp[-(\log\tau - \log\tau_g)^2 / 2\log^2\mu] \qquad (2a)$$

where τ is the aerosol optical depth, τ_g is the geometric mean and log μ is the geometric standard deviation (see Aitchison and Brown, (1972) for example). Throughout the text "log" refers to $\log_{10}$ while "ln" represents $\log_e$. The representation of the ℓ distribution in τ space is given by;

$$P_\ell(\tau) = \frac{1}{\tau}\frac{1}{\ln 10}P_\ell(\log\tau) \qquad (2b)$$

This functional representation is mathematically inconsistent (one does not just substitute the argument "τ" for "log τ" in $P_\ell(\log\tau)$) but we have retained the formulation to keep the nomenclature as simple as possible. The n distribution in τ space is;

$$P_n(\tau) = \frac{1}{\sqrt{2\pi}\,\sigma(\tau)_n}\exp[-(\tau - <\tau>_n)^2 / 2\sigma(\tau)_n{}^2] \qquad (3)$$

Table 1 summarizes the nomenclature of these three analytical distributions while Figure 1 illustrates the form of the distributions and their basic statistical parameters. When a particular parameter is computed for a given analytical

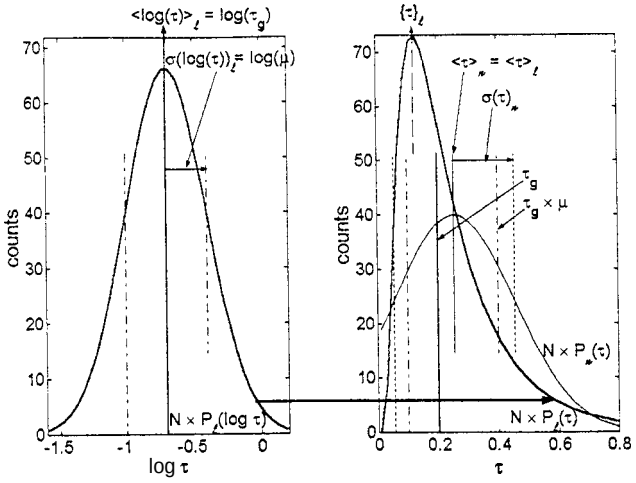

Figure 1; the left hand and right hand panels show the probability distributions and associated parameters in log τ and τ space respectively (c.f. Table 1). The arithmetic mean of the normal distribution is set equal to the arithmetic mean of the lognormal representation in linear τ space (i.e. $<\tau>_n = <\tau>_\ell$).

1. GSFC/NASA, Greenbelt, MD
2. NOAA/NESDIS/Climate Research and Applications Division. Camp Springs, MD
3. L.O.A. from CARTEL, Université de Sherbrooke, Sherbrooke, Québec, Canada

Paper number 2000GL011581.
0094-8276/00/2000GL011581$05.00

Table 1. Nomenclature for the analytical distributions. The analytical expressions under the $P_\ell(\tau)$ column are derived from equation (2b). See Fig. 1 for illustrations of most of these parameters.

	log τ space	τ space	
	$P_\ell(\log\tau)$	$P_\ell(\tau)$	$P_n(\tau)$
mean	$\langle\log\tau\rangle_\ell \rightarrow \tau_g = 10^{\langle\log\tau\rangle_\ell}$	$\langle\tau\rangle_\ell = \tau_g e^{\ln^2\mu/2}$	$\langle\tau\rangle_n$
stan. dev.	$\sigma(\log\tau)_\ell \rightarrow \mu = 10^{\sigma(\log\tau)_\ell}$	$\sigma(\tau)_\ell = \langle\tau\rangle_\ell \sqrt{e^{\ln^2\mu} - 1}$	$\sigma(\tau)_n$
peak	$\{\log\tau\}_\ell$	$\{\tau\}_\ell = \tau_g e^{-\ln^2\mu}$	$\{\tau\}_n$
median	$[\log\tau]_\ell$	$[\tau]_\ell = \tau_g$	$[\tau]_n$

distribution the symbol for the distribution is given as a subscript; no subscript indicates that the statistical parameter is computed for a data histogram (see Figure 2 below).

The geometric standard deviation μ is key to reporting the geometric mean and its variance in a form which is as intuitive as the arithmetic mean and its standard deviation;

$$\tau_g \times \mu^{\pm 1} \tag{4}$$

The symmetrical limits corresponding to this formulation in log τ space and the associated asymmetric limits in τ space are shown as dash-dot lines in the left hand and right hand panels of Figure 1 respectively.

Below we compare and evaluate the quality of ℓ and n fits to data histograms. These fits will not be in the sense of minimum residuals but rather in the more pragmatic sense of allowing the analytical frequency distribution ($N \times P_\ell(\log\tau)$ or $N \times P_n(\tau)$) to assume the same mean, standard deviation and number of measurements as the data histogram.

Some tests for the quality of normal or lognormal representations of data histograms

Two common higher order parameters for the characterization of data histograms are the skewness and kurtosis (γ_1 and γ_2 defined in Ambramowitz and Stegun [1972]). Skewness is an indicator of

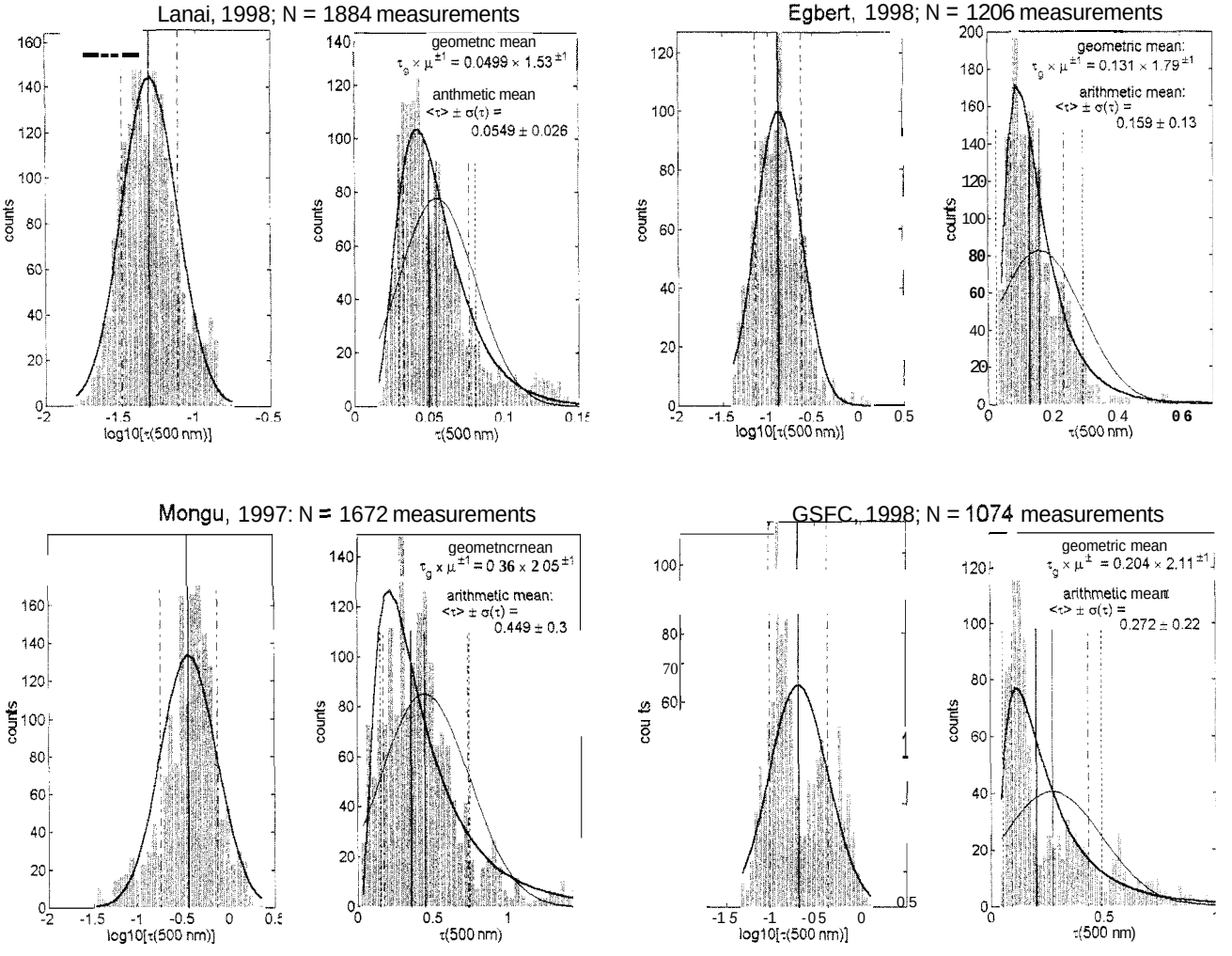

Figure 2; sample histograms for selected cases. Figure 1 illustrates the meaning of the different curves and vertical lines (the same boldness and line types are retained in Figure 2). (a) examples for which the ℓ fit is clearly superior to the n fit (b) fits with problems related to negative skewness or bi-modality in log τ space.

Table 2. Station and data ensemble parameters. Last two columns show associated aerosol classes for each station.

station	long.	lat.	ASL (m.)	land cover	data years available	aerosol background class	aerosol influence(s) on top of background
Waskesiu, Sask., CAN	N53°55'	W 106'04'	550	boreal forest	96 - 98	rural	biomass burning
GSFC. MD, USA	N39°01'	W76°52'	50	suburban	96 - 98	industrial & rural	urban emissions, maritime aerosols
Egbert, Ont., **CAN**	N44° 13'	W79°45'	264	farmland	98	industrial & rural	urban emissions
Thompson, Man.. CAN	N55°47'	W97°50'	218	boreal forest	96 - 99	rural	biomass burning
Bondville, IL, USA	N40°03'	W88°22'	212	farmland	96, 98, 99	rural	industrial
Lanai, HI, USA	N20°49'	W156°59'	80	island	97, 98	maritime	volcanic emissions / Asian dust
Mongu. Zambia	S15°15'	E23°09'	1107	savanna	96 - 98	rural	biomass burning
Bahrain	N26°19'	E50°30'	0	island	98	maritime/dust	industrial

distribution asymmetry and is negative for a distribution displaying a left hand tail, positive for a right hand tailed distribution and zero for a normal distribution. Kurtosis is an indicator of the peakedness of a distribution and is postive for a very peaked distribution, negative for a flat distribution and zero for a normal distribution.

Skewness and kurtosis are measures of the general form of the data histogram and can be used as higher order indicators of how closely the form resembles a normal distribution. A test which permits a more intuitive understanding of the quality of an n or ℓ fit to a data histogram is to ascertain whether the fitted curve correctly predicts the AOD position of histogram features other than those used in constraining the fit. One such test is to estimate the histogram peak position (mode) in τ space which, as will be seen in the data histogram examples below, is not co-located with the mean.

Another feature-position test is to ascertain whether the representation of the ℓ distribution in τ space can be used to predict the arithmetic mean (in the case of the n distribution the test is irrelevant since the n distribution mean is set equal to the

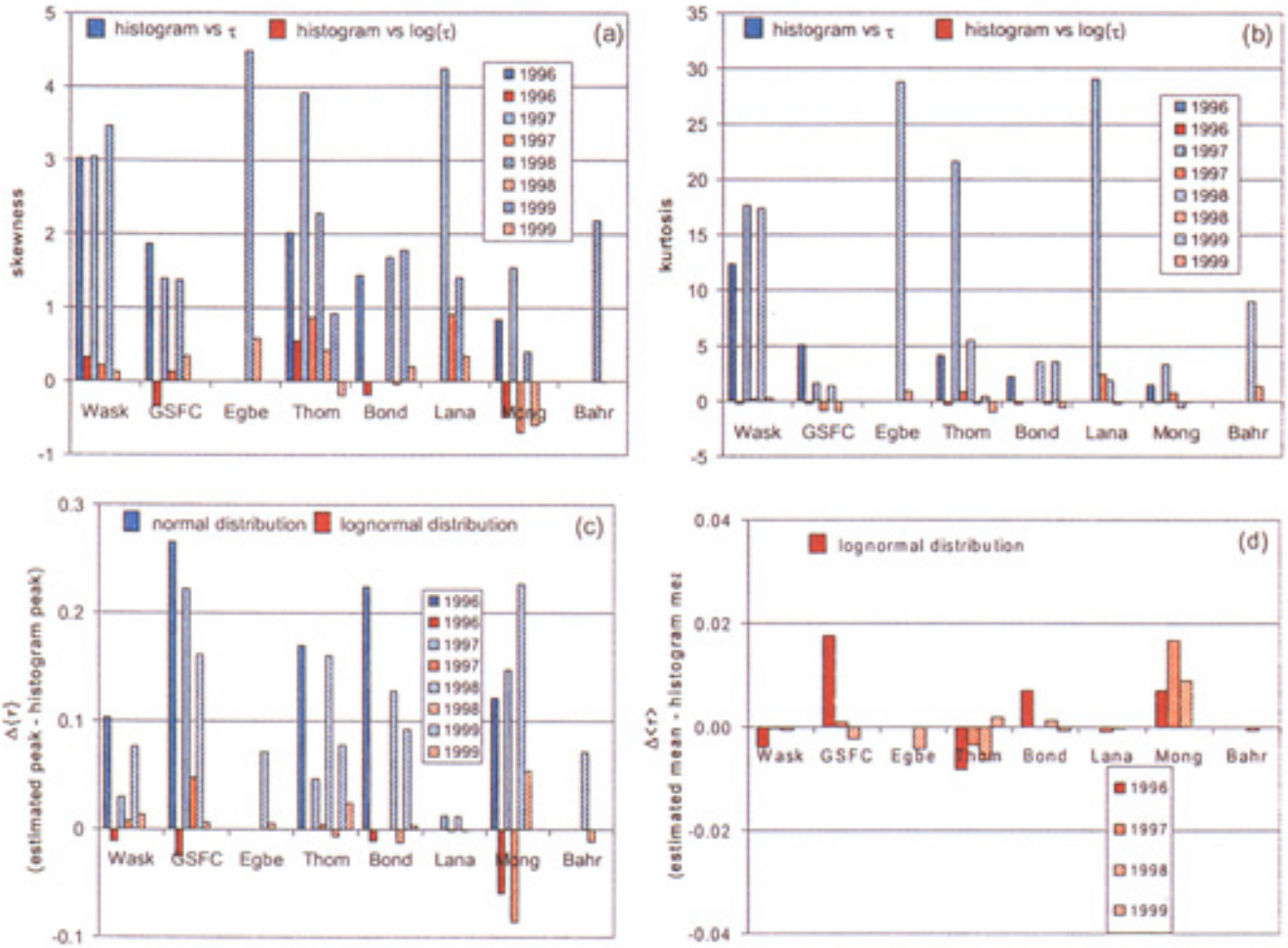

Figure 3; (a) histogram skewness (skewness for a normal distribution, whether in log τ space or τ space, is zero). (b) histogram kurtosis (kurtosis for a normal distribution is zero). (c) error in the estimated histogram peak position in τ space as estimated using the lognormal fit distribution. (d) error in the arithmetic mean of the histogram as estimated using the lognormal fit distribution.

arithmetic mean of the histogram). This arithmetic mean estimation test is interesting from the standpoint that a generally small mean error would ensure continuity with data sets of AOD arithmetic mean simply by employing the basic statistical parameters of the ℓ distribution to compute the arithmetic mean.

Logarithmic and linear representations of optical depth histograms

in this section we present a sampling of histograms using data acquired by CIMEL sunphotometers of the AERONET network over a variety of stations and a variable number of years. Detailed specifications of the AERONET instruments and data acquisition system are described elsewhere (Holben et al., 1998).

Table 2 is a listing of the stations from which data were acquired and those years for which data was available. The table includes information on the background regional aerosol as well as major aerosol influences which may dominate local sunphotometry at a given station. The choice of stations was largely influenced by a desire to represent the greatest possible variety of aerosol types. The study was limited to a standard wavelength of 500 nm.

All the data chosen in our study were cloud screened according to the procedure defined in Smirnov et al. (2000). Three months of data was taken as the standard sampling period in order to achieve a frequency of measurements which was of sufficient density to permit a significant number of measurements per sampling bin. These three months corresponded to the summer period of June, July and August except in the case of Mongu where the August to October period was chosen in order maximize the influence of the biomass burning season. Thirty sampling bins between extreme AOD values in log τ space and sixty sampling bins between extreme AOD values in τ space were used in the generation of data histograms. This number of bins seemed to give reasonably smooth histograms and distribution fits in both spaces; however it was ascertained that the generated statistical parameters were fairly insensitive to bin number and bin width.

Figure 2a shows some selected sample histograms along with the ℓ and n fits in log τ and τ space. The two-panel figures were designed so that Figure I could serve as a reference template to indicate the salient features of the ℓ and n analytical distributions respectively. These figures also provide an indication of the variation in the geometric mean and geometric standard deviation μ. The analytical fits qualitatively demonstrate the superiority of the ℓ fit over the n fit both in log τ space where the histogram is generally more symmetric and normal in appearance and in linear τ space where the asymmetric form of the ℓ representation is clearly better matched to the positively skewed form of the data histograms.

Figure 2b shows some sample histograms where the ℓ fits were still generally superior to the n fits but where certain features in the log τ histogram distribution degraded the quality of the ℓ fit. These include the bi-modal features in the GSFC histogram of 1998 and the negative skewness in the Mongu histogram of 1997.

Figure 3 shows the four test parameters of skewness, kurtosis, peak location error and arithmetic mean estimation error for all stations and available years. Figures 3a and 3b demonstrate that the skewness and kurtosis calculated in log τ space is systematically more normal like (closer to zero) than the equivalent calculations in τ space. The τ space histograms are positively skewed; although the log τ space representation shows some positive skewness it is significantly less than the former.

Figures 3c and 3d show the errors in histogram peak location and estimated arithmetic mean. For all stations and all years the rms average peak location error for the n distribution fit was 0.13 while the rms average error for the ℓ representation was 0.03. The use of the ℓ representation to estimate the arithmetic mean and standard deviation of the histogram in τ space yielded rms errors of 0.01 and 0.04 respectively for all stations and all years (the arithmetic mean for the n distribution fit is set equal to the histogram arithmetic mean as indicated above). Thus the ℓ distribution can be used to estimate histogram features in linear τ space to accuracies which are of the order of or a little greater than typical sunphotometry errors of 0.01 to 0.02.

Conclusion

Multi-year and multi-station AOD data was employed to demonstrate that the lognormal probability distribution was systematically a better reference for reporting AOD statistics than a normal probability distribution. Comparative tests in log τ and τ space showed that data histograms in the former nearly always corresponded to smaller values of skewness and kurtosis and accordingly that this was a better space for a normal representation of the histograms. The estimation of the AOD value corresponding to the histogram peak in τ space was significantly better if a lognormal fit was applied to the data histogram. The use of the lognormal distribution to predict the arithmetic mean and standard deviation of the histogram in τ space yielded reasonably accurate estimates and as such provides a means of ensuring the continuity of data archives based on arithmetic means. in certain cases the data histograms displayed apparent bi-modal features or negative skewness in log τ space which neither distribution could adequately fit but for which the lognormal distribution was still a better reference.

Acknowledgements

The authors thank NASA and the National Science Foundation (NRC) for their support. The thoughtful advice of Sasha Smirnov is gratefully acknowledged.

References

Aitchison, J.. Brown, J. A. C., The lognormal distribution, *Cambridge University Press,* 176 pp., 1957.

Ambramowitz, M., Stegun. 1. A., Handbook of Mathematical Functions, Dover Publications Inc.. New York, 1972.

Campbell, J. W., The lognormal distribution as a model for bio-optical variability in the sea, *J. Geophys. Res.*, Vol. 100, No. C7, pp. 13237-13254, 1995.

Malm, W. C., Walther, E. G., Cudney, R. A., The Effects of Water Vapor, Ozone and Aerosol on Atmospheric Turbidity, *Jour. App. Met.* Vol. 16, pp. 268-274, 1977.

King, M. D., Byrne, D. M., Reagan, J. A., Herman, B. M., Spectral Variation of Optical Depth at Tucson Arizona between August 1975 and December 1977, *Jour. App. Met.,* Vol. 19, pp. 723-732, 1980.

Holhen, B.N., T.F.Eck, I.Slutsker, D.Tanre, J.P.Buis, A.Setzer, E.Vermote, J.A.Reagan, Y.J.Kaufman, T.Nakajima, F.Lavenu, I.Jankowiak, and A.Smirnov, AERONET - A federated instrument network and data archive for aerosol characterization, *Rem.Sens.Env.*, *66(1)*, 1-16, 1998.

Ignatov, A, Stowe, L., Physical Basis, Premises, and Self-Consistency Checks of Aerosol Retrievals from TRMM/VIRS, submitted to the *Jour. App. Met.,* 1999.

Ignatov, A, Yu. Ratner, personal communication, 1995.

Smimov, A., Holben, B. N., Eck, T. F., Dubovik, O., Slutsker, I., Cloud Screening and Quality Control Algorithms for the AERONET data base, accepted for publication in *Remote Sens. Environ.*, 2000.

(Received March 13, 2000; revised August 3, 2000;
accepted August 7, 2000.)